On the Importance of Visualisation in a Data Driven Society

Daniel Archambault[a]
Newcastle University, U.K.

Keywords: Visualisation.

Abstract: Machine learning and data science are receiving significant attention and rightly so. The results that can be produced by distilling large amounts data are amazing. Yet, what society expects is human oversight at an appropriate level and trust of system results. Oversight and trust are not for machines; oversight and trust are for humans. Effective solutions thus require careful human-machine collaboration and in turn careful visualisation design. This paper motivates why visualisation design forms a necessary part of a data driven society. It provides motivation for carefully designed visualisations that must take into account human perceptual factors, target audience, and the automated processes applied to the information before visualisation. It provides practical examples where all three must be carefully thought about in order to deliver effective data science.

1 INTRODUCTION

Data science and artificial intelligence have made remarkable strides over the past decade with results far exceeding what we would have expected. These methods process large amounts of data to achieve these results. In a modern day, data-driven society, the data, results of the models, and insights into how the models work need to be made accessible to diverse user communities so that they can be informed and make data-driven decisions.

Data science is often defined as extracting knowledge and insights from data. Machine learning and data science methods are essential to solve this problem. Without them, it would not be possible to process data at scale. However, the human side of the data science equation is often forgotten.

Insight is not for computers;
Insight is for humans.

Machine learning and data science methods are the engine whereas visualisation is an equally important windscreen of a data driven society. Yes, it is true that we would not get anywhere without scalable data processing methods and models, but the final step of human consumption is often not emphasised enough.

Governments and societies see the value of (and fear) the capabilities of artificial intelligence. They seek oversight of the models and processes to construct these models and apply them to the large data sets. Again, this is a human-centred problem aligned with visualisation and not one aligned with machines.

Oversight of models and AI is not for computers;
Oversight of models and AI is for humans.

As these concepts require the human, methods to expose the models and data from the machine are a requisite part of the solution. Thus, visualisation is critical for effective artificial intelligence, machine learning, and data science. Visualisation is necessary for insight into data (and therefore an essential part of data science) and is required for oversight of models and AI. In this paper, we explore this question from a variety of perspectives and my personal experience.

2 IS A TABLE OF STATISTICS THE ANSWER?

Is it the case that a table of statistics is sufficient for this problem? Perhaps the simplest form of visualisation is the best? The importance of visualisations complementing calculations has been made clear many decades ago in statistics (Anscombe, 1973):

> *A computer should make <u>both</u> calculations <u>and</u> graphs. Both sorts of output should be studied; each will contribute to understanding.*

More recent work further illustrates this point. The Datasaurus Dozen (Matejka and Fitzmaurice,

[a] https://orcid.org/0000-0003-4978-8479

Archambault, D.
On the Importance of Visualisation in a Data Driven Society.
DOI: 10.5220/0013445500003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 3: VISAPP*, pages 7-10
ISBN: 978-989-758-728-3; ISSN: 2184-4321

2017) shows that many very different plots can be generated for the same set of statistics. Many different patterns – even a dinosaur – could be present in your data for the same statistics. A similar idea can be seen in the area of graph drawing (Chen et al., 2021; Chen et al., 2018) making graph drawings essential and complimentary to graph mining and graph statistics. Although statistics are important, they need to be complemented by effective and relevant visualisations for insight.

Also, more studies should be run on the perceptual effectiveness of visualisations. The visual channel used can influence the amount of order seen in the data (Chung et al., 2016). Small multiple representations of dynamic graphs are often faster, with equivalent accuracy, for dynamic graph visualisations (Archambault et al., 2011). The benefits of keeping a stable drawing, also known as preservation of the mental map, is surprisingly very difficult to quantify (Archambault and Purchase, 2016), but there is evidence that it helps identify specific nodes and specific paths through a graph in a visually scalable way (Archambault and Purchase, 2013).

Thus, complimentary and relevant visualisations are needed with the proper encoding.

3 CAN DATA PROCESSING AND VISUALISATION BE APPLIED INDEPENDENTLY?

Another possibility is that any visualisation can be applied independently from the data processing algorithm. Unfortunately, this is not the case either as the data processing influences how the visualisations are perceived. Therefore, the automatic processing must be selected with the visualisation tasks in mind. Ether the viewer will need to know the limitations of the automatic processing algorithm or the processing algorithm needs to be selected so the chances of misinterpretation is minimised.

Graph sampling methods take a large graph and generate a representative sample for visualisation or other processing purposes. If the end goal is visualisation, it is important to understand that the sampling method chosen influences how the sample is perceived. For example, relative degree is important for perceiving nodes as high degree in a sample (Wu et al., 2017). In terms of edge bundling, often edges are bundled together based on similar edge properties but this can create the appearance of false connections and patters in the data where there are none (Lhuillier et al., 2017). If the idea is to group edges together with similar origins and destinations together then this visualisation is suitable. However, if underlying paths and topology is required for understanding, edge bundling techniques that group edges with underlying paths (Wallinger et al., 2022; Wallinger et al., 2023) would be more suitable which can be computed while simultaneously drawing the graph (Archambault et al., 2024).

In summary, the data processing algorithm also can influence the visualisation. This should be selected with the task and user background in mind so that the chances of a misleading visualisation are minimised.

4 CASE STUDY: VISUALISING MODELS FOR DIVERSE AUDIENCES DURING COVID-19

Visualisation was embedded into the UK response to COVID-19 through RAMPVis (Chen et al., 2022; Dykes et al., 2022): a volunteer project to support visualisation of pandemic modelling for a variety of audiences. The pandemic response was a concrete example of what data scientists do. Data science methods, models, and visualisation were visible on a daily basis in media across the world. During this time, new strategies for developing visualisations were created for a wide range of audiences.

In our case, we were tasked on visualisation of simulations of policies on contact tracing data. In the early stages of the pandemic, there was no time for careful selection of problems or design of bespoke solutions. Also, the communities that we were working with did not know that visualisation was a field and how it could help. Fortunately, we did have some experience with visualising contact tracing that we could draw upon (Baumgartl et al., 2021), but not for the specific problem of simulating policies on contact tracing simulation and understanding how these models performed. In our initial sessions, existing visualisation tools were applied out of the box (Simonetto et al., 2020; Arleo et al., 2022) to the simulations and presented to pandemic modelling teams to discover what was necessary (Sondag et al., 2022). These sessions provided buy in for visualisations as initial results could be seen and errors in the simulations could be discovered. For example, visualisations produced from MulitDynNoS (Arleo et al., 2022) were able to illustrate that too many random cases (imported from outside) were set in the simulator when travel was already restricted. As the situation evolved, more bespoke tools could be developed using standard visualisation methodologies. This resulted in a final visualisation (Sondag et al., 2022) that could be used to make sense of simulated contact tracing policies.

In a second project, we were tasked with the idea of communicating pandemic modelling ideas to children through interactive visualisations (Mcneill et al., 2023). Working with Technocamps, we developed an interactive visualisation method to help children explore disease modelling concepts to gain a better understanding what data scientists would do during a pandemic. The visualisation was used by 100 young people every month at its height, providing a real data science problem that was relevant to their experiences.

5 CONCLUSION

In summary, both insight and oversight are not for computers; they are for humans. A way to achieve communication of data a models inside a computer to humans is through visualisation. Thus, visualisation forms an equally important part of data science and AI. This has been known for a long time in statistics, but sometimes requires repeating. These visualisations cannot be applied blindly without considering the underlying algorithms used to process the data as it influences how the visualisation is perceived. A case study on how this work was applied during the COVID-19 response was discussed along with human-centred development strategies and considerations for communicating data science concepts to diverse audiences.

ACKNOWLEDGEMENTS

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC-BY) licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

Anscombe, F. (1973). Graphs in statistical analysis. *The American Statistician*, 27(1):17–21.

Archambault, D., Liotta, G., Nöllenburg, M., Piselli, T., Tappini, A., and Wallinger, M. (2024). Bundling-Aware Graph Drawing. In *32nd International Symposium on Graph Drawing and Network Visualization (GD 2024)*, volume 320, pages 15:1–15:19.

Archambault, D., Purchase, H., and Pinaud, B. (2011). Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics*, 17(4):539–552.

Archambault, D. and Purchase, H. C. (2013). Mental map preservation helps user orientation in dynamic graphs. In *Graph Drawing (GD '12)*, pages 475–486.

Archambault, D. and Purchase, H. C. (2016). Can animation support the visualisation of dynamic graphs? *Information Sciences*, 330:495–509.

Arleo, A., Miksch, S., and Archambault, D. (2022). Event-based dynamic graph drawing without the agonizing pain. *Computer Graphics Forum*, 41(6):226–244.

Baumgartl, T., Petzold, M., Wunderlich, M., Hohn, M., Archambault, D., Lieser, M., Dalpke, A., Scheithauer, S., Marschollek, M., Eichel, V. M., Mutters, N. T., Consortium, H., and Landesberger, T. V. (2021). In search of patient zero: Visual analytics of pathogen transmission pathways in hospitals. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):711–721.

Chen, H., Soni, U., Lu, Y., Huroyan, V., Maciejewski, R., and Kobourov, S. (2021). Same stats, different graphs: Exploring the space of graphs in terms of graph properties. *IEEE Transactions on Visualization and Computer Graphics*, 27(3):2056–2072.

Chen, H., Soni, U., Lu, Y., Maciejewski, R., and Kobourov, S. (2018). Same stats, different graphs. In *Graph Drawing and Network Visualization (GD '18)*, pages 463–477.

Chen, M., Abdul-Rahman, A., Archambault, D., Dykes, J., Ritsos, P., Slingsby, A., Torsney-Weir, T., Turkay, C., Bach, B., Borgo, R., et al. (2022). RAMPVIS: Answering the challenges of building visualisation capabilities for large-scale emergency responses. *Epidemics*, 39:100569.

Chung, D. H. S., Archambault, D., Borgo, R., Edwards, D. J., Laramee, R. S., and Chen, M. (2016). How ordered is it? on the perceptual orderability of visual channels. *Computer Graphics Forum*, 35(3):131–140.

Dykes, J., Abdul-Rahman, A., Archambault, D., Bach, B., Borgo, R., Chen, M., Enright, J., Fang, H., Firat, E. E., Freeman, E., et al. (2022). Visualization for epidemiological modelling: challenges, solutions, reflections and recommendations. *Philosophical Transactions of the Royal Society A*, 380(2233):20210299.

Lhuillier, A., Hurter, C., and Telea, A. (2017). State of the art in edge and trail bundling techniques. *Computer Graphics Forum*, 36(3):619–645.

Matejka, J. and Fitzmaurice, G. (2017). Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 1290–1294.

Mcneill, G., Sondag, M., Powell, S., Asplin, P., Turkay, C., Moller, F., and Archambault, D. (2023). From asymptomatics to zombies: Visualization-based education of disease modeling for children. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23.

Simonetto, P., Archambault, D., and Kobourov, S. (2020). Event-based dynamic graph visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 26(7):2373–2386.

Sondag, M., Turkay, C., Xu, K., Matthews, L., Mohr, S., and Archambault, D. (2022). Visual analytics of contact tracing policy simulations during an emergency response. *Computer Graphics Forum*, 41(3):29–41.

Wallinger, M., Archambault, D., Auber, D., Nollenburg, M., and Peltonen, J. (2022). Edge-Path Bundling: A Less Ambiguous Edge Bundling Approach . *IEEE Transactions on Visualization & Computer Graphics*, 28(1):313–323.

Wallinger, M., Archambault, D., Auber, D., Nöllenburg, M., and Peltonen, J. (2023). Faster edge-path bundling through graph spanners. *Computer Graphics Forum*, 42(6):e14789.

Wu, Y., Cao, N., Archambault, D., Shen, Q., Qu, H., and Cui, W. (2017). Evaluation of graph sampling: A visualization perspective. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):401–410.